

# Mathematical treatment of the canonical finite state machine for the Ising model: $\epsilon$ -machine

E. Rodríguez Orta and E. Estevez-Rams\*

*Facultad de Física-Instituto de Ciencias y Tecnología de Materiales,  
University of Havana (IMRE), San Lazaro y L. CP 10400. La Habana. Cuba.*

R. Lora Serrano

*Universidade Federal de Uberlandia, AV. Joao Naves de Avila,  
2121- Campus Santa Monica, CEP 38408-144, Minas Gerais, Brasil.*

(Dated: April 4, 2016)

The complete framework for the  $\epsilon$ -machine construction of the one dimensional Ising model is presented correcting previous mistakes on the subject. The approach follows the known treatment of the Ising model as a Markov random field, where usually the local characteristic are obtained from the stochastic matrix, the problem at hand needs the inverse relation, or how to obtain the stochastic matrix from the local characteristics, which are given via the transfer matrix treatment. The obtained expressions allow to perform complexity-entropy analysis of particular instance of the Ising model. Three examples are discussed: the 1/2-spin nearest neighbor and next nearest neighbor Ising model, and the persistent biased random walk.

## I. INTRODUCTION

$\epsilon$ -machine reconstruction or computational mechanics is an approach to discover the nature of patterns and to quantify it [1, 2]. Building from information theory concepts, it has found applications in several fields and proved its value in a number of context[3–5]. Its use in statistical mechanics allows to define and calculate magnitudes that complements thermodynamic quantities such as entropy, specific heat, correlation length or structure factor[6, 7]. For a stochastic process considered to be stationary, the  $\epsilon$ -machine is its optimal minimal description, understood, as having the best (most accurate) predictive power while using the least possible resources (minimal statistical complexity)[8]. In computational mechanics, the notion of causality is taken in a general temporal sense, in a given context, cause to effect relations are established between past to future events[9].

Although a general framework, computational mechanics has been thoroughly developed for discrete time and space process by several authors [3, 9–11].

One of the earlier developments in computational mechanics has been the analysis of patterns in one dimensional spin systems under Ising-type models[6, 7, 12, 13]. The use of the  $\epsilon$ -machine reconstruction for the one dimensional Ising model allowed the analysis of such systems as a deterministic finite state machine, and close expressions for the entropy density, statistical complexity and excess entropy were found[6, 7, 12]. Unfortunately, this treatments failed to understand that the Ising model has to be considered a Markov (Gibbs) random field in order to correctly determine the probability measure of the associated Markov process[14]. The failure of not considering correctly the Markov random field, leads to incorrect expressions for the stochastic matrix governing the conditional probability between spin blocks. As a consequence, all quantitative results derived from such matrix are incorrect and this includes the determination of the causal states, the statistical complexity, the entropy density and the excess entropy. This includes the worked example on the Ising 1/2-nearest neighbor model and next nearest neighbor model.

In this contribution the framework of computational mechanics for the Ising model from the most general setting in one dimension is developed. It is shown how to obtain the stochastic matrix from the Markov field: an inverse problem not usually treated on the literature. The results goes beyond correcting the mathematical treatment. An important consequence of the presented mathematical development is whether is it possible to cast the Ising model, in the general case, a single emission spin sequential process. We finally show the use of the developed framework via three examples.

The paper is organized as follows. In section II the transfer matrix method for the Ising model is shortly reviewed for self-completeness and to fix the used notation. In section III we describe the Ising model as a random field and deduce the expression for the stochastic conditional matrix. In section IV, building from the previous sections, the

---

\*Electronic address: estevez@imre.oc.uh.cu

$\epsilon$ -machine formalism for the Ising model is discussed and the entropic magnitudes are introduced. Finally, in section V worked examples applying the developed mathematical tools is presented. Conclusions follow.

## II. THE TRANSFER MATRIX FORMALISM

The Ising (-Lenz) model is probably the most studied lattice type model in statistical mechanics, and is well covered in several statistical physics books for nearest 1/2 spin neighbor interactions[15]. Let us briefly recap, for completeness and notation purposes, the basic ideas of the transfer matrix formalism (we closely follow Dobson[16]) but in a general setting of a local type interaction Hamiltonian which is usually not found in texts.

Consider an one dimensional chain of discrete values of length  $L$ :

$$s^L = s_0 s_1 s_2 \dots s_{L-1},$$

where  $s_i$  can take values from a finite alphabet  $\Theta$  of cardinality  $\theta(=|\Theta|)$  (there will be  $\theta^L$  possible sequences  $s^L$ ). Each individual  $s_i$  is called a spin. The interaction energy between spins of the sequence has a finite range  $n$  such that it can be written as

$$E(s_i, s_{i\pm k}) = \begin{cases} \Lambda(s_i, s_{i\pm k}) & 0 < k \leq n \\ 0 & k > n \end{cases}. \quad (1)$$

The  $s^L$  sequence can be partitioned into blocks of length  $n$ , considering  $L = Nn$ ,

$$s^L = [s_0 s_1 \dots s_{n-1}] [s_n s_{n+1} \dots s_{2n-1}] \dots [s_{(N-1)n} s_{(N-1)n+1} \dots s_{Nn-1}],$$

which can be relabeled

$$\begin{aligned} s^L &= [s_0^{(0)} s_1^{(0)} \dots s_{n-1}^{(0)}] [s_0^{(1)} s_1^{(1)} \dots s_{n-1}^{(1)}] \dots [s_0^{(N-1)} s_1^{(N-1)} \dots s_{n-1}^{(N-1)}]. \\ &= \eta_0 \eta_1 \dots \eta_{N-1}. \end{aligned}$$

Where

$$\eta_i = s_0^{(i)} s_1^{(i)} \dots s_{n-1}^{(i)}. \quad (2)$$

The set of all possible blocks  $\eta_i$  will be denoted by  $\Upsilon$  with cardinality  $v = \theta^n$ .  $\Upsilon$  will be taken as an ordered set (e.g lexicographic order) and to each  $\eta_i$  a natural number, between 0 and  $v - 1$ , will be assigned. In what follows,  $\eta_i$  should be understood not only as the configuration (2) but also as its corresponding order in the set  $\Upsilon$ , context will eliminate any ambiguity.

As the interaction energy has range  $n$ , one spin corresponding to the  $\eta_i$  block, can only interact with all the spins within  $\eta_i$  (type I) and at least one spin from the adjacent blocks  $\eta_{i\pm 1}$  (type II).

Taking into account that  $\Lambda(s_i, s_j) = \Lambda(s_j, s_i)$ , the interaction energy of type I for the  $\eta_p$  block, in the presence of an external field B, will be

$$x_{\eta_p} = -B \sum_{i=0}^{n-1} s_i^{(p)} + \sum_{i=0}^{n-2} \sum_{k=i+1}^{n-1} \Lambda(s_i^{(p)}, s_k^{(p)}), \quad (3)$$

which defines a vector  $\langle X |$  of length  $v$ . The contribution of type II will be denoted by  $y_{\eta_p \eta_{p+1}}$ , and will be given by

$$y_{\eta_p \eta_{p+1}} = \sum_{i=0}^{n-1} \sum_{k=0}^i \Lambda(s_i^{(p)}, s_k^{(p+1)}), \quad (4)$$

which defines a  $v \times v$  matrix. In general  $Y_{\eta_i \eta_j} \neq Y_{\eta_j \eta_i}$  which makes  $Y$  non symmetric. The energy of the whole configuration  $s^L$  can then be written as

$$\Lambda(s^L) = x_{\eta_0} + y_{\eta_0 \eta_1} + x_{\eta_1} + y_{\eta_1 \eta_2} + \dots + y_{\eta_{N-2} \eta_{N-1}} + x_{\eta_{N-1}}. \quad (5)$$

The vector  $\langle U |$  and the matrix  $V$  are then introduced as

$$u_{\eta_i} = \exp(-\frac{1}{2}\beta x_{\eta_i}) \quad (6)$$

$$v_{\eta_i \eta_j} = \exp[-\beta(\frac{1}{2}x_{\eta_i} + y_{\eta_i \eta_j} + \frac{1}{2}x_{\eta_j})]. \quad (7)$$

where  $\beta \equiv (k_B T)^{-1}$  is the Boltzmann product.  $V$  is known as the transfer matrix.

The partition function follows

$$\begin{aligned} Z_{Nn} &= \sum_{\eta_0=0}^{v-1} \sum_{\eta_1=0}^{v-1} \dots \sum_{\eta_{N-1}=0}^{v-1} \exp[-\beta \Lambda(s^L)] \\ &= \langle U | V^{N-1} | U \rangle \end{aligned} \quad (8)$$

for open (free) boundary conditions. For close (periodic) boundary conditions

$$Z_{Nn} = \text{Tr}(V^N). \quad (9)$$

$\text{Tr}(M)$  denotes the trace of the matrix  $M$ .

As the trace of a matrix is invariant to similarity transformations, from equation (9) for close boundary conditions

$$Z_{Nn} = \sum \lambda_i^N \quad (10)$$

follows.  $\lambda_i$  are the eigenvalues of the matrix  $V$ . If  $\lambda_i$  is degenerate, then the term is added as much times as its multiplicity. If the eigenvalues are labeled in non increasing order ( $|\lambda_i| \geq |\lambda_j| \rightarrow j > i$ ), then for  $N \gg 1$

$$Z_{Nn} = \lambda_0^N \quad (11)$$

where  $\lambda_0$  is known as the dominant eigenvalue and, according to the Perron-Frobenius theorem, it is real, positive and non-degenerate[14].

For open boundary conditions, using again the Perron- Frobenius theorem for a square positive defined matrix  $V$ , the following holds

$$\lim_{N \rightarrow \infty} \frac{V^N}{\lambda_0^N} = |{}_r a_0\rangle \langle {}_l a_0|, \quad (12)$$

where  $\langle {}_l a_0 |$  y  $|{}_r a_0\rangle$  are respectively, the left and right eigenvectors corresponding to the dominant eigenvalue. The eigenvectors are normalized in the sense of  $\langle {}_l a_0 | {}_r a_0 \rangle = 1$ . The matrix  $|{}_r a_0\rangle \langle {}_l a_0|$  is known as Perron projection matrix. Using (12) and (8) we arrive at

$$Z_{Nn} = \langle U | {}_r a_0 \rangle \langle {}_l a_0 | U \rangle \lambda_0^{N-1} \quad (13)$$

which in the particular case of a diagonalizable matrix reduces to

$$Z_{Nn} = u_0^2 \langle a_0 | a_0 \rangle \lambda_0^{N-1} \quad (14)$$

and  $u_i$  are the components of the vector  $\langle U |$  in the orthogonal base defined by the eigenvectors  $\langle a_i |$ . It is well documented how the thermodynamic magnitudes can be obtained from the partition function [15].

### III. ISING MODEL AS A MARKOV (GIBBS) RANDOM FIELD

The probability of a given spin chain will be given by

$$\begin{aligned} \text{Pr}(s^L) &= \frac{1}{Z_{Nn}} e^{-\beta \Lambda(s^L)} \\ &= \frac{1}{Z_{Nn}} (U_{\eta_0} V_{\eta_0 \eta_1} V_{\eta_1 \eta_2} \dots V_{\eta_{N-2} \eta_{N-1}} U_{\eta_{N-1}}) \\ &= \frac{U_{\eta_0} U_{\eta_{N-1}}}{M \lambda_0^{N-1}} \prod_{i=0}^{N-2} V_{\eta_i \eta_{i+1}} \end{aligned} \quad (15)$$

valid for free boundary conditions, and where we have written  $M \equiv \langle U | r a_0 \rangle \langle l a_0 | U \rangle$  [21]. For periodic boundary conditions

$$Pr(s^L) = \frac{1}{\lambda_0^N} \prod_{i=0}^{N-2} V_{\eta_i \eta_{i+1}} \quad (16)$$

asked in both cases  $N \gg 1$ .

Consider the spin chain divided in two half, the first half will be called the past

$$\overleftarrow{s}^L = s_{-L} s_{-L+1} \dots s_{-1} = \eta_{-N} \eta_{-N+1} \dots \eta_{-1},$$

while the second half, called the future, is given by

$$\overrightarrow{s}^L = s_0 s_1 \dots s_{L-1} = \eta_0 \eta_1 \dots \eta_{N-1}.$$

The spin  $s_0$  is sometimes referred to as the present.

The known relation

$$Pr(\overrightarrow{s}^L | \overleftarrow{s}^L) = \frac{Pr(\overleftarrow{s}^L, \overrightarrow{s}^L)}{Pr(\overleftarrow{s}^L)} \quad (17)$$

may tempt to use equations (15) or (16), but one must be warned against doing so. Both expressions must be understood as valid for whole systems (spin chains) and not for portions of the systems. In this way,  $Pr(\overleftarrow{s}^L)$  in the denominator of (17) can not be taken as the probability of the isolated configuration  $\overleftarrow{s}^L$  instead, use must be made of the relation

$$Pr(\overleftarrow{s}^L) = \sum_{\{\overrightarrow{s}^L\}} Pr(\overleftarrow{s}^L, \overrightarrow{s}^L). \quad (18)$$

The reader can check that the result is not the same. This is at the heart of the failure in the mathematical treatment in [[6, 7, 12]] the random field character of the system is overlooked.

The Ising model is a particular case of a Gibbs random field[14]. In a formal way, a set of sites  $\Psi = 0, 1, 2, \dots, N-1$  is given, together with a finite set  $\Upsilon$  of cardinality  $v$ . There is a correspondence  $x : \Psi \rightarrow \Upsilon$  such that  $s^L$  represents the configuration where each site  $i \in \Psi$  has been assigned a value  $\eta_i \in \Upsilon$ .  $\Psi^\Upsilon$  is the set of all possible configurations. Consider a probabilistic measure  $\mathbb{P}$  associated to the space  $\Psi^\Upsilon$ .  $\{\Psi^\Upsilon, \mathbb{P}\}$  will be called a random field and  $\eta_i$  will be random variables with stochastic dependence among them. For each  $\eta_i \in \Upsilon$  the conditional probabilities

$$\begin{aligned} Pr(\eta_i = \eta | \eta_t = b_t \text{ for } t \in \Psi, t \neq i, b_t \in \Upsilon) \\ = Pr(\eta_i = \eta | s^L - \eta_i), \end{aligned} \quad (19)$$

are well defined, where  $s^L - \eta_i$  denotes the configuration  $s^L$  excluding site  $i$  (a better notation would have been  $s^L - i$ , but the introduced one is sufficiently clear and convenient in what follows).

The case of interest is one where it is not necessary to know the whole configuration  $s^L - \eta_i$  to determine the probabilities (19) but it suffices to know the values  $\eta_t$  in a subset  $\mathcal{N}_i \subset \Psi$ , which is called neighborhood and complies with the conditions

1. For  $i \in \Psi$ ,  $\mathcal{N}_i$  is a subset (possibly empty) of  $\Psi$  that does not contain  $i$ .
2.  $t \in \mathcal{N}_i \rightarrow i \in \mathcal{N}_t$

A Markov random field is defined as one fulfilling

$$Pr(\eta_i = \eta | s^L - \eta_i) = Pr(\eta_i = \eta | \eta_t = b_t \text{ for } t \in \mathcal{N}_i, b_t \in \Upsilon). \quad (20)$$

These probabilities are called the local characteristics associated to  $\mathbb{P}$ .

Let  $\mathcal{N}_k = k-1, k+1$  for  $k = 1, 2, \dots, N-2$ ,  $\mathcal{N}_0 = 1$ ,  $\mathcal{N}_{N-1} = N-2$ . Then, with respect to this system of neighborhood of the Markov random field, there is associated a Markov chain  $\{x_i\}_{i=0, \dots, N-1}$ .

If  $s^L - \eta_i$  is the configuration  $s^L$  without considering the block  $\eta_i$ , then, the probability  $Pr(\eta_i | s^L - \eta_i)$  that the  $i$ -block has value  $\eta_i$  when all the other spins (that is, excluding the  $\eta_i$  block) will have the configuration  $s^L - \eta_i$  will be given by the product rule

$$Pr(\eta_i | s^L - \eta_i) = \frac{Pr(s^L)}{Pr(s^L - \eta_i)} = \frac{Pr(s^L)}{\sum_{s^{L*}} Pr(s^{L*})} \quad (21)$$

where the sum  $s^{L*}$  is over all configurations identical to  $s^L$  except, possibly, for the block  $\eta_i$ .

Using equation (15), the probability of a configuration will be

$$\begin{aligned}
 Pr(s^L) &= \frac{1}{Z_{Nn}} e^{-\beta \Lambda(s^L)} \\
 &= \frac{1}{Z_{Nn}} \exp \left[ -\beta \left( \sum_{i=0}^{N-1} x_{\eta_i} + \sum_{i=0}^{N-2} y_{\eta_i \eta_{i+1}} \right) \right] \\
 &= \frac{1}{Z_{Nn}} e^{-\beta x_{\eta_{N-1}}} \prod_{j=1}^{N-2} e^{-\beta x_{\eta_j}} e^{-\beta y_{\eta_j \eta_{j+1}}}
 \end{aligned} \tag{22}$$

and,

$$\begin{aligned}
 Pr(s^L - \eta_i) &= \sum_{s^{L*}} Pr(s^{L*}) = \\
 &= \left[ \frac{1}{Z_{Nn}} e^{-\beta x_{\eta_{N-1}}} \prod_{\substack{j=1 \\ j \neq i; j \neq i-1}}^{N-2} e^{-\beta x_{\eta_j}} e^{-\beta y_{\eta_j \eta_{j+1}}} \right] e^{-\beta x_{\eta_{i-1}}} \\
 &\sum_{\eta_k} e^{-\beta y_{\eta_{i-1} \eta_k}} e^{-\beta x_{\eta_k}} e^{-\beta y_{\eta_k \eta_{i+1}}}.
 \end{aligned} \tag{23}$$

For the local characteristics equation (21) will now be

$$\begin{aligned}
 Pr(\eta_i | s^L - \eta_i) &= \frac{e^{-\beta y_{\eta_{i-1} \eta_i}} e^{-\beta x_{\eta_i}} e^{-\beta y_{\eta_i \eta_{i+1}}}}{\sum_{\eta_k} e^{-\beta y_{\eta_{i-1} \eta_k}} e^{-\beta x_{\eta_k}} e^{-\beta y_{\eta_k \eta_{i+1}}}} \\
 &= \frac{V_{\eta_{i-1} \eta_i} V_{\eta_i \eta_{i+1}}}{\sum_{\eta_k} V_{\eta_{i-1} \eta_k} V_{\eta_k \eta_{i+1}}}
 \end{aligned} \tag{24}$$

for blocks  $\eta_i$  not at the extremes.

For the first block

$$\begin{aligned}
 Pr_0(\eta_0 | s^L - \eta_i) &= \frac{e^{-\beta x_{\eta_0}} e^{-\beta y_{\eta_0 \eta_1}}}{\sum_{\eta_k} e^{-\beta x_{\eta_k}} e^{-\beta y_{\eta_k \eta_1}}} \\
 &= \frac{U_{\eta_0} V_{\eta_0 \eta_1}}{\sum_{\eta_k} U_{\eta_k} V_{\eta_k \eta_1}}
 \end{aligned} \tag{25}$$

Similar expression can be found for the last block. Expression (24) has the important consequence that

$$Pr(\eta_i | s^L - \eta_i) = Pr(\eta_i | \eta_{i-1}, \eta_{i+1}) \tag{26}$$

It must be noted that one of the consequence of equations (24) and (25), is that the probability of  $\eta_i$  taking a particular value is conditional on the values of the immediate neighboring blocks, and not by the whole configuration  $s^L - \eta_i$ , this determines the Markov character of the random field.

The stochastic matrix  $P$  will be defined by

$$P_{ij} = Pr(\eta_j | \eta_i). \tag{27}$$

which defines a transition probability from state  $\eta_i$  to state  $\eta_j$ . By definition  $\sum_j P_{ij} = 1$ . If  $\langle p^\infty |$  is the vector of probabilities over the blocks  $\eta_i$  then is well known that the stationary distribution[14] is given by

$$\langle p^\infty | = \langle w_0 | \tag{28}$$

where  $\langle w_0 |$  is the left dominant eigenvector of the matrix  $P$ . The vector  $\langle p^\infty |$  allows to calculate  $Pr(\eta_i)$  when the Markov process has been running for a sufficiently long time.

The local characteristics can be written in terms of the stochastic matrix  $P$  using Bayes theorem

$$\begin{aligned}
 Pr(\eta_i, \eta_{i-1}, \eta_{i+1}) &= Pr(\eta_i | \eta_{i-1}, \eta_{i+1}) Pr(\eta_{i+1}, \eta_{i-1}) = \\
 &= Pr(\eta_i | \eta_{i-1}, \eta_{i+1}) Pr(\eta_{i+1} | \eta_{i-1}) Pr(\eta_{i-1})
 \end{aligned} \tag{29}$$

similarly

$$\begin{aligned} Pr(\eta_i, \eta_{i-1}, \eta_{i+1}) &= Pr(\eta_{i+1} | \eta_i, \eta_{i-1}) Pr(\eta_i, \eta_{i-1}) = \\ &= Pr(\eta_{i+1} | \eta_i, \eta_{i-1}) Pr(\eta_i | \eta_{i-1}) Pr(\eta_{i-1}) \end{aligned} \quad (30)$$

now equating both (29) and (30)

$$\begin{aligned} Pr(\eta_i | \eta_{i-1}, \eta_{i+1}) &= \frac{Pr(\eta_i | \eta_{i-1}) Pr(\eta_{i+1} | \eta_i, \eta_{i-1})}{Pr(\eta_{i+1} | \eta_{i-1})} \\ &= \frac{Pr(\eta_i | \eta_{i-1}) Pr(\eta_{i+1} | \eta_i)}{\sum_l Pr(\eta_l | \eta_{i-1}) Pr(\eta_{i+1} | \eta_l)} \end{aligned} \quad (31)$$

where the Markov character of the field has been used and the last step is justified by the total probability theorem[22].

Equation (31) can be rewritten as

$$Pr(\eta_i | \eta_{i-1}, \eta_{i+1}) \sum_l Pr(\eta_l | \eta_{i-1}) Pr(\eta_{i+1} | \eta_l) = Pr(\eta_i | \eta_{i-1}) Pr(\eta_{i+1} | \eta_i) \quad (32)$$

which forms, when written for each  $\eta_i$ , an homogeneous system of quadratic forms. Such system can have non trivial solution if it is undetermined, which happens in this case if the square of the number of unknown is larger than the number of equations.

As each local characteristic is determined by three  $\eta$ 's, there will be  $\theta^{3n} = \nu^3$  equations and  $\theta^{2n} = \nu^2$  unknowns. The relations

$$\sum_i^\nu Pr(\eta_i | \eta_j) = 1 \quad \forall j$$

must be added that eliminates  $\nu$  unknowns. The total number of unknowns is then  $\nu(\nu - 1)$  and the total number of equations  $\nu^3$ .

$$\nu^2(\nu - 1) < \nu^2(\nu - 1)^2 \implies \nu > 2.$$

For  $\nu = 2$ , the system will also have non trivial solution as a result of the additional reduction of equations by the symmetry of the transfer matrix. Returning to equation (32) and rewriting for any local characteristic

$$Pr(\eta_i | \eta_j, \eta_m) \sum_l Pr(\eta_l | \eta_j) Pr(\eta_m | \eta_l) = Pr(\eta_i | \eta_j) Pr(\eta_m | \eta_i), \quad (33)$$

introducing

$$Y_l(j, m) = Pr(\eta_l | \eta_j) Pr(\eta_m | \eta_l) \quad (34)$$

equation (33) can be written as

$$Pr(\eta_i | \eta_j, \eta_m) \sum_l Y_l(j, m) = Y_i(j, m). \quad (35)$$

The normalization condition ( which can be derived from equation (31)) over the local characteristics determines

$$\sum_k^\nu Pr(\eta_k | \eta_j, \eta_m) = 1, \quad (36)$$

Equation (35) is linear and homogeneous over the  $Y_i(j, m)$  which, upon solving for the non trivial-case, leads to a system of simple homogeneous quadratic equations (34) which can be readily solved.

The Markov character of the system means that the generation process can forget of all the past except the last block  $\eta_{-1}$  (the last  $n$  spins), to determine, as certain as possible the future. In other words, if the local characteristics implies a stochastic matrix as equation (33) imply, then all past configuration  $\overleftarrow{s}^L$  with the same last block  $\eta_{-1}$  conditions (statistically) the same future, this fact allows to consider the Ising chain as a canonical finite state machine known as  $\epsilon$ -Machine.

#### IV. ISING MODEL AS A CANONICAL FINITE STATE MACHINE: $\epsilon$ -MACHINE

Causal states are at the core of the computational mechanic framework[9]. If two blocks  $\eta_{-1}$  and  $\eta'_{-1}$  give the same  $Pr(\vec{s}^L | \overleftarrow{s}^L)$ , for all possible futures  $\vec{s}^L$ , then  $\eta_{-1}$  y  $\eta'_{-1}$  are said to belong to the same causal state ( $C_p$ ) and we write  $\eta_{-1} \sim \eta'_{-1}$ ,  $\eta_{-1}, \eta'_{-1} \in C_p$ . Two blocks belonging to the same causal state  $C_p$  define identical rows in the stochastic matrix.

The partition of the set  $\Upsilon$  in classes of causal states is an equivalence relation complying with the transitivity condition (if  $\eta_i \sim \eta_j$  y  $\eta_j \sim \eta_k$  then  $\eta_i \sim \eta_k$ ), symmetry (if  $\eta_i \sim \eta_j$  then  $\eta_j \sim \eta_i$ ) and reflectivity ( $\eta_i \sim \eta_i$ ). The set of causal state uniquely determines the future of a sequence. The set of all causal states will be denoted by  $\mathcal{C}$  with cardinality  $|\mathcal{C}|$ . A function  $\epsilon$  can be defined over  $\eta$ ,  $\epsilon : \Upsilon \rightarrow \mathcal{C}$  which relates  $\eta$  with its causal states  $C$ ,

$$\epsilon(\eta) \equiv C = \{\eta' | Pr(\vec{s} | \eta) = Pr(\vec{s} | \eta') \quad \forall \vec{s}\}.$$

The probability of a (recursive) causal state is directly deducible from equation (28),

$$Pr(C_p) = \sum_{\eta_j \in C_p} Pr(\eta_j) = \sum_{\eta_j \in C_p} p_{\eta_j}^\infty \quad (37)$$

As each causal state represents the set of past that determines (probabilistically) the same future, the set of causal state represents the memory the system has to keep in order to predict, as good as possible, the future.

The statistical complexity has been defined as the Shannon entropy over the causal states[8]

$$\begin{aligned} C_\mu &\equiv - \sum_{C_p \in \mathcal{C}} Pr(C_p) \log Pr(C_p) \\ &= H[\mathcal{C}]. \end{aligned} \quad (38)$$

The logarithm is usually taken in base two and the units are then bits. Statistical complexity is a measure of how much memory (resources) the system needs to optimally predict the future. If the system has  $|\mathcal{C}|$  causal states then the statistical complexity has the upper bound

$$C_\mu \leq \log |\mathcal{C}|,$$

corresponding to a uniform distribution of probabilities over the causal states. The relation follows from a known property of the Shannon entropy. The upper bound of the statistical complexity is also known as topological entropy. As already stated, for a fixed number of causal states, the increase of statistical complexity is the result of a more uniform distribution over the causal state, in this sense such increase witness the increase of uncertainty over the occurrence of such states and therefore the unreductible entropy of the system (e.g. thermal noise). The system needs more resources to account for the unpredictable behavior of the system.

The probability of occurrence of block  $\eta_i$  conditional on the causal state  $C$  will be given by

$$\begin{aligned} Pr(\eta_i | C) &= \sum_{\eta_k} Pr(\eta_i | \eta_k \cap C) Pr(\eta_k | C) \\ &= Pr(\eta_i | \eta_j; \eta_j \in C) \sum_{\eta_k \in C} Pr(\eta_k | C) \\ &= Pr(\eta_i | \eta_j; \eta_j \in C) \end{aligned} \quad (39)$$

in the first step the total probability theorem was used, in the second step use has been made of the fact that conditioning in  $\eta_k \cap C$  is equal to conditioning in  $\eta_k$  if the block belongs to the causal state  $C$  and, finally  $\sum_{\eta_k \in C} Pr(\eta_k | C) = 1$ .

$$\begin{aligned} Pr(\eta_j) &= \sum_{\eta_k \in \Sigma} Pr(\eta_j | \eta_k) Pr(\eta_k) \\ &= \sum_{C_k \in \mathcal{C}} \sum_{\eta_k \in C_k} Pr(\eta_j | \eta_k) Pr(\eta_k) \\ &= \sum_{C_k \in \mathcal{C}} Pr(\eta_j | \eta_{k'}, \eta_{k'} \in C_k) \sum_{\eta_k \in C_k} Pr(\eta_k) \\ &= \sum_{C_k \in \mathcal{C}} Pr(\eta_j | \eta_{k'}, \eta_{k'} \in C_k) Pr(C_k) \end{aligned}$$

now we make use of equation (39) to get

$$Pr(\eta_j) = \sum_{C_k \in \mathcal{C}} Pr(\eta_j | C_k) Pr(C_k). \quad (40)$$

which allows to compute the occurrence of a block from the probabilities over the causal states.

The probability of a transition from one causal state to other will be given by

$$Pr(C_k \rightarrow C_p) \equiv Pr(C_p|C_k) = \sum_{\eta_j \in C_p} Pr(\eta_j|C_k) \quad (41)$$

If we define the transition matrix  $T^{(\eta)}$ , whose elements are the probability of going from state  $C_k$  to state  $C_p$  upon emitting a block  $\eta$ :

$$T_{rq}^{(\eta)} = Pr(C_r \xrightarrow{\eta} C_q). \quad (42)$$

By construction, the emission of a block  $\eta$  determines uniquely the causal state to where the transition occurs (this is called the unifilar property[9]). In this sense, the generation process is deterministic. Correspondingly, the connectivity matrix  $T$  is defined as

$$T_{rq} = \sum_{\eta \in \Sigma} T_{rq}^{(\eta)} = Pr(C_r \rightarrow C_q) \quad (43)$$

which connects causal states without regard of the emitted block.

An  $\epsilon$ -machine is defined by the causal state function  $\epsilon$ , which relates histories with causal states and the transition matrices between them. In the case of the Ising model, the  $\epsilon$ -machine is represented by a deterministic finite state machine with vertex defined by the causal states and edges labeled according to the transition probabilities[12].

In order to account for the irreducible randomness, the density of entropy is defined as[13]

$$\begin{aligned} h_\mu &= \lim_{L \rightarrow \infty} H[\eta_0 | \eta_{-L} \eta_{-L+1} \dots \eta_{-1}] \\ &= \lim_{L \rightarrow \infty} H[\eta_0 | \overleftarrow{s}^L] \end{aligned} \quad (44)$$

$h_\mu$  is the uncertainty on the next emitted block  $\eta_0$  conditional on having seen infinite previous blocks (spins). By definition  $h_\mu \geq 0$ .

$$\begin{aligned} H[\eta_0 | \overleftarrow{s}^L] &= - \sum_{\eta_0} \sum_{\eta_{-n}} \dots \sum_{\eta_{-1}} Pr(\overleftarrow{s}^L \eta_0) \log Pr(\eta_0 | \overleftarrow{s}^L) \\ &= - \sum_{\eta_0} \sum_{\eta_{-N}} \dots \sum_{\eta_{-1}} Pr(\eta_0 | \overleftarrow{s}^L) Pr(\overleftarrow{s}^L) \log Pr(\eta_0 | \overleftarrow{s}^L) \end{aligned} \quad (45)$$

where  $\overleftarrow{s}^L = \eta_{-N} \eta_{-N+1} \dots \eta_{-1}$ , , on the other hand, using  $\overleftarrow{s}^{L-n} = \eta_{-N} \dots \eta_{-2}$

$$\begin{aligned} Pr(s^L) &= Pr(\eta_{-1} | s^{L-n}) P(s^{L-n}) \\ &= Pr(s^{L-n} | \eta_{-1}) Pr(\eta_{-1}). \end{aligned} \quad (46)$$

In the last step, use has been made of Bayes theorem. Substituting equation (46) on equation (45) and reordering terms

$$\begin{aligned} H[\eta_0 | \overleftarrow{s}^L] &= - \sum_{\eta_0} \left[ \sum_{\eta_{-1}} Pr(\eta_0 | \eta_{-1}) Pr(\eta_{-1}) \log Pr(\eta_0 | \eta_{-1}) \right. \\ &\quad \left. \left\{ \sum_{\eta_{-N}} \dots \sum_{\eta_{-2}} Pr(\overleftarrow{s}^{L-n} | \eta_{-1}) \right\} \right], \end{aligned} \quad (47)$$

now

$$\sum_{\eta_{-N}} \dots \sum_{\eta_{-2}} Pr(\overleftarrow{s}^{L-n} | \eta_{-1})$$

is the probability that from  $\eta_{-1}$  any configuration is conditioned and that probability is 1. Equation (47) then reduces to

$$\begin{aligned} h_\mu &= \lim_{L \rightarrow \infty} H[\eta_0 | \overleftarrow{s}^L] = H[\eta_0 | \eta_{-1}] \\ &= - \sum_{\eta_j \in \Sigma} Pr(\eta_j) \sum_{\eta_i \in \Sigma} Pr(\eta_i | \eta_j) \log Pr(\eta_i | \eta_j) \\ &= - \sum_{C_\alpha \in \mathcal{C}} Pr(C_\alpha) \sum_{\eta_k \in \Sigma} Pr(\eta_k | C_\alpha) \log Pr(\eta_k | C_\alpha) \end{aligned} \quad (48)$$



The mutual information between past and future is called the excess entropy[13]

$$E_\mu \equiv I[\overleftarrow{s} : \overrightarrow{s}], \quad (49)$$

where  $I[X : Y]$  is the mutual information between  $X$  e  $Y$ . From the finite range character of the interaction in the Ising model

$$\begin{aligned} E_\mu &= I[\overleftarrow{s} : \overrightarrow{s}] = I[\eta_{-1} : \eta_0] \\ &= H[\eta_{-1}] - H[\eta_0 | \eta_{-1}] \end{aligned} \quad (50)$$

where

$$H[\eta_i] = \sum_{\eta_i \in \Sigma} Pr(\eta_i) \log Pr(\eta_i), \quad (51)$$

and

$$H[\eta_0 | \eta_{-1}] = H[\eta_0 | \overleftarrow{s}^L] = h_\mu$$

given by equation (48).

From equation (38) and (48) we arrive to the expression

$$E_\mu = C_\mu - h_\mu. \quad (52)$$

Excess entropy is a measure of the resources needed by the system in order to optimally predict the future, once the irreducible randomness has been subtracted[13]. As  $E$  is a mutual information, it will always be a non-negative value, which implies

$$C_\mu \geq h_\mu$$

If the system is completely ordered then  $h_\mu = 0$  and

$$C_\mu = E_\mu.$$

#### A. Markov process as emission of single spins

The result, that in order to correctly derive the stochastic matrix, one needs the local characteristic of the Markov field, points to fact that is not straight forward to cast a spacially extended process into a sequential process. In the Markov field, all spin values over the whole one dimensional lattice is given simultaneously. In general, the spin value at one site depends on the spin values to the right, as well as to the left, within the interaction range. It is the transfer matrix method that rewrites the whole system as a one over blocks of size equal to the interaction range, that allows to cast the dynamics of the system as Markov process, but in doing so, the building blocks cease to be the individual spins and instead the  $\eta$  blocks are the new individual entities. It is then, not straight forward to consider valid, if one can describe the  $\epsilon$ -machine process over single spin emission. This is contrast with previous treatment of the subject [6, 12] and some general conclusion derived from it [17]. To exemplify the problems associated with defining the Markov process over the emission of a single spin, consider a general Hamiltonian with interaction range  $n = 2$ , in such system a eight period perfect sequence is possible with probability larger than zero. The four state FSM with single spin emission is incapable of topologically reproducing such sequence. There is then the need to introduce “bogus” states additional to the  $\eta$  blocks. Yet, the stochastic matrix is well defined by the local characteristic over the  $\eta$  blocks and the  $\epsilon$ -machine over block emission makes the introduction of unnatural states unnecessary.

Having said so, in what follows the Markov process will be formally cast in terms of a single spin process to bridge the gap with the mathematical developments made before. In the previous treatment, the Markov process has been constructed taking as a single event the emission of a whole block  $\eta$  of length  $n$ . Other treatments have considered the Markov process single event the emission of a single spin[6, 7, 12]. Let us introduce the operator  $\pi$  such that if  $\eta_k = s_k s_{k+1} \dots s_{k+n-2} s_{k+n-1}$ , then

$$\pi \eta_k = s_{k+1} \dots s_{k+n-1}$$

and  $\pi \eta_k$  has length  $n - 1$ . Similarly

$$\eta_k \pi = s_k \dots s_{k+n-2}$$

The probability of emitting a single spin  $s_0 = s$  conditioned in being on the causal state  $C$  will be denoted by  $Pr(s|C)$ . Using equation (39),

$$\begin{aligned} Pr(s|C) &= \sum_{s_1} \sum_{s_2} \dots \sum_{s_{n-1}} Pr(\eta_k|\eta_j; \eta_j \in C) \\ &= \sum_{\substack{\eta_k \in \mathcal{Y} \\ s_0 = s}} Pr(\eta_k|\eta_j; \eta_j \in C) \end{aligned} \quad (53)$$

and the sum is over all configurations  $\eta_i$  with spin  $s_0 = s$ . It is important to notice that upon emitting a spin  $s$  a transition occurs from  $\overleftarrow{s}^L$  to  $\overleftarrow{s}^L s$  which implies a new causal state to which

$$\eta_0 = \pi\eta_{-1}s = s_{-n+1} \dots s_{-1}(s_0 = s)$$

belongs and which is uniquely determined by the generated spin. The transition  $\eta_j \rightarrow \pi\eta_j s$  conditional on the emission of spin  $s$ ,  $Pr(\pi\eta_j s|s, \eta_j)$ , is equal to 1 if the transition is allowed.

If  $\eta_i$  and  $\eta_j$  are taken to belong to the same causal state  $C$ , then by construction

$$Pr(\overrightarrow{S} \in sF|\eta_i) = Pr(\overrightarrow{S} \in sF|\eta_j), \quad (54)$$

where  $F$  is the set of all possible futures and  $sF$  is the operation of prefixing to each future the spin  $s$ . It is clear that  $\overrightarrow{S} \in sF$  can be split into events:  $s_0 = s$  and  $\pi\overrightarrow{S} \in F$ , then

$$\begin{aligned} Pr(\overrightarrow{S} \in sF|\eta_i) &= Pr(\overrightarrow{S} \in sF|\eta_j) \\ Pr(s_0 = s, \pi\overrightarrow{S} \in F|\eta_i) &= Pr(s_0 = s, \pi\overrightarrow{S} \in F|\eta_j). \end{aligned} \quad (55)$$

Using the product rule on both sides and taking into account that for three random variables  $X, Y$  y  $Z$ ,  $Pr(Z \in A, Y = y|X = x) = Pr(Z \in A|Y = y, X = x)Pr(Y = y|X = x)$ ,

$$\begin{aligned} Pr(\pi\overrightarrow{S} \in F|s_0 = s, \eta_i)Pr(s_0 = s|\eta_i) &= \\ = Pr(\pi\overrightarrow{S} \in F|s_0 = s, \eta_j)Pr(s_0 = s|\eta_j), \end{aligned} \quad (56)$$

and because  $Pr(s_0 = s|\eta_i) = Pr(s_0 = s|\eta_j)$  is defined strictly positive,

$$\begin{aligned} Pr(\pi\overrightarrow{S} \in F|s_0 = s, \eta_i) &= Pr(\pi\overrightarrow{S} \in F|s_0 = s, \eta_j) \\ Pr(\pi\overrightarrow{S} \in F|\pi\eta_i s) &= Pr(\pi\overrightarrow{S} \in F|\pi\eta_j s) \end{aligned} \quad (57)$$

Expression (57) is equivalent to say that  $\pi\eta_i s$  and  $\pi\eta_j s$  belong to the same causal state and therefore, the emission of one spin determines uniquely the causal state transition.

Therefore, the following holds

$$Pr(C_k \xrightarrow{s} C_i) = Pr(s|C_k).$$

Now, the transition matrix  $T^{(s)}$  whose elements are the probability of making a transition from state  $C_k$  to state  $C_p$  upon emitting a spin  $s$ , will be:

$$T_{rq}^{(s)} = Pr(C_r \xrightarrow{s} C_q) = Pr(s|C_r), \quad (58)$$

and  $\pi\eta_r s \in C_q$ ,  $\eta_r \in C_r$ . The transition matrix between causal states regardless of the emitted symbol will be given by

$$T_{rq}^{(s)} = \sum_{\substack{\pi\eta_r s \in C_q \\ \eta_r \in C_r}} Pr(s|C_r), \quad (59)$$

where the left dominant eigenvector determines the probability of the causal state (equation (28) and (37)). The statistical complexity  $C'_\mu$  follows.

The entropy density will now be

$$h'_\mu = \lim_{L \rightarrow \infty} H[s|s_{-L}s_{-L+1} \dots s_{-1}] \quad (60)$$

$h'_\mu$  is now the uncertainty in the next emitted spin  $s$  conditional of having observed infinite preceding spins in the past. Now

$$\begin{aligned} h'_\mu &= \lim_{L \rightarrow \infty} H[s|\overleftarrow{s}^L] = H[s|\eta_{-1}] \\ &= -\sum_{C_\alpha \in \mathcal{C}} Pr(C_\alpha) \sum_{s \in \Theta} Pr(s|C_\alpha) \log Pr(s|C_\alpha) \end{aligned} \quad (61)$$

The excess entropy will now be

$$E'_\mu = C'_\mu - nh'_\mu, \quad (62)$$

$n$  is the interaction range (the size of the  $\eta$  blocks).

## V. EXAMPLES

### A. 1/2 nearest neighbors spin chain

The 1/2-spin Ising model for the nearest neighbor interaction is defined by the interaction Hamiltonian[18]

$$E = -B \sum_i s_i - J \sum_j s_j s_{j+1}, \quad (63)$$

where  $B$  is the external field, and  $J$  is the interaction parameter. Both parameters are independent. The  $\eta$  blocks set will be

$$\eta = \downarrow, \uparrow.$$

The local characteristics derived from equation (24) reduce to

$$\begin{aligned} Pr(\downarrow | \downarrow, \downarrow) &= \frac{e^{4\beta J}}{e^{2\beta B} + e^{4\beta J}} \\ Pr(\downarrow | \downarrow, \uparrow) &= \frac{1}{e^{2\beta B} + 1} \\ Pr(\uparrow | \downarrow, \downarrow) &= \frac{1}{e^{4\beta J - 2\beta B} + 1} \\ Pr(\uparrow | \downarrow, \uparrow) &= \frac{e^{2\beta B}}{e^{2\beta B} + 1} \\ Pr(\downarrow | \uparrow, \downarrow) &= \frac{1}{e^{2\beta B} + 1} \\ Pr(\downarrow | \uparrow, \uparrow) &= \frac{1}{e^{2\beta(B+2J)} + 1} \\ Pr(\uparrow | \uparrow, \downarrow) &= \frac{e^{2\beta B}}{e^{2\beta B} + 1} \\ Pr(\uparrow | \uparrow, \uparrow) &= \frac{e^{2\beta(B+2J)}}{e^{2\beta(B+2J)} + 1}. \end{aligned} \quad (64)$$

Solving the linear system of equation (35), results in the system of quadratic equations

$$\begin{aligned} Pr(\downarrow | \uparrow) Pr(\uparrow | \downarrow) &= e^{2\beta B - 4\beta J} Pr(\downarrow | \downarrow) \\ Pr(\uparrow | \uparrow) &= e^{2\beta B} Pr(\downarrow | \downarrow), \end{aligned} \quad (65)$$

together with the normalization conditions

$$\begin{aligned} Pr(\downarrow | \uparrow) + Pr(\uparrow | \uparrow) &= 1 \\ Pr(\downarrow | \downarrow) + Pr(\uparrow | \downarrow) &= 1 \end{aligned} \quad (66)$$

lead to the solution for[23]

$$Pr(\uparrow | \uparrow) = \frac{2e^{2\beta J}}{\sqrt{4e^{2\beta B} + e^{4\beta(B+J)} - 2e^{2\beta(B+2J)} + e^{4\beta J} + e^{2\beta(B+J)} + e^{2\beta J}}}. \quad (67)$$

And the other entries of the stochastic matrix follows.

From the stochastic matrix, the two state  $\epsilon$ -machine was built and  $(h_\mu, E)$  were calculated for  $10^5$  points taking randomly the value of the parameters. The corresponding plot is shown in figure 1. This type of complexity map has been discussed before[13]. Small values of disorder can accommodate a large range of excess entropy values, which means varying probability between the two possible causal states. As disorder increases, the system loses structure tending towards a single state process which although increasingly random is also increasingly less complex.

### B. 1/2 next nearest neighbors spin chain

The 1/2-spin Ising model for the next nearest neighbor interaction is defined by the interaction Hamiltonian[18]

$$E = -B \sum_i s_i - J_1 \sum_j s_j s_{j+1} - J_2 \sum_k s_k s_{k+2}. \quad (68)$$

where  $B$  is the external field, and  $J_1, J_2$  are spin coupling parameters. The  $\eta$  blocks set will be

$$\eta = \downarrow\downarrow, \downarrow\uparrow, \uparrow\downarrow, \uparrow\uparrow$$

we can take the value 1 to correspond to spin up, whereas  $-1$  corresponds to spin down.

Figure 2 shows the excess entropy as a function of  $J_1$  and  $J_2$  for the ground state ( $\beta \rightarrow \infty$ ) at zero field ( $B = 0$ ) and as a function of  $J_1/B$  and  $J_2/B$  for non-zero field ( $B \neq 0$ ). The excess entropy together with the  $\epsilon$ -machine reconstruction allows to distinguish four orderings of periodicity 1,2,3 and 4, identified as the ordered sequences

- (1)  $\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow \dots$
- (2)  $\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow \dots$
- (3)  $\uparrow\uparrow\downarrow\uparrow\uparrow\downarrow \dots$
- (4)  $\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow \dots$

which are the expected phases[18]. The pure ferromagnetic phase (1) and pure anti-ferromagnetic phase (2) are constructed from a single causal state, while (4) needs two causal states. The ordered phase (3) requires three causal states and it only happens for  $B \neq 0$ .

### C. Persistent biased one dimensional random walk

Consider a one dimensional random walk in which at each event the walker moves one step in the same direction as the previous event with probability  $p$  (the probability of changing the direction of movement is then  $1 - p$ ), we also define the probability that the random walker will move to the right (left) at each event by  $r$  ( $1 - r$ ). The described model is known as the persistent biased random walk[19, 20] and has been found to be isomorphous with the nearest neighbor Ising chain. The developed formalism has been used to calculate the excess entropy as a function of the probabilities  $r$  and  $p$ , as well as the entropy density, both are shown in figure 3. The  $\epsilon$ -machine reconstruction for different value range of  $r$  and  $p$  is shown in figure 4. There should be a symmetry in  $r$  and  $1 - r$ , as there is no distinction between “right” and “left”, and indeed this is the case, so only values of  $r$  up to  $1/2$  are shown. For arbitrary values of the probabilities the finite state machines has two causal states (figure 4d) and disorder is a consequence of the competing “interaction” between (anti)persistence of motion direction whose strength is given by  $p$ , and the movement in some fixed direction measured by  $r$ .

If  $r$  is small, and  $p$  is near one, then persistence takes charge and the random walk will be mostly to the left with seldom changes of direction, this is seen in the low values of  $h_\mu$ . At  $p = 1$ , persistence overtakes any other behavior, and the  $\epsilon$ -machine is a single causal state equivalent to a ferromagnetic state (figure 4a). If  $r$  is small ( $\approx 0$ ) and  $p$  is also small ( $\approx 0$ ), the random walker has a strong tendency to change the movement direction at every step, which is

the absolute dominant behavior at  $p = 0$  resulting in a perfect anti-ferromagnetic state (figure 4b). At  $p = r = 1/2$  the maximum state of disorder is reached which is equivalent to an unbiased coin tossing (figure 4c). In the entropy plot (Fig 3a) a large central region of almost no structure is seen, this same region is where large value of disorder is achieved as seen in the  $h_\mu$  plot (Fig. 3b).

Finally, a complexity-entropy map[13] of the model is shown in figure 3c. Compare the map with that of the 1/2-nearest neighbor Ising model.

## VI. CONCLUSION

$\epsilon$ -machines reconstruction or computational mechanics, is a powerful tool in the analysis of complexity, which has been used in a wealth of different theoretical and practical situations. Unfortunately, for the case of the one dimensional Ising model, previous treatment were flawed by improper treatment of the model as a Markov random field, which, it gave wrong numerical results and misleading behaviors for specific realization of the model. In this paper we developed the complete formalism for  $\epsilon$ -machine construction of the one dimensional Ising model. The given expressions can be then used to model specific instance of the interaction Hamiltonian, opening the way to the correct use of computational mechanics for the analysis of complexity behavior in the one-dimensional Ising model. A computer library with the full implementation of the framework can be requested from the authors.

## Acknowledgments

This work was partially financed by FAPEMIG under the project BPV-00047-13. EER which to thank PVE/CAPES for financial support under the grant 1149-14-8. Infrastructure support was given under project FAPEMIG APQ-02256-12.

- 
- [1] J.P. Crutchfield, in *Modeling complex phenomena*, ed. by L. Lam, V. Naroditsty (Springer, Berlin, 1992), pp. 66–101
  - [2] J.P. Crutchfield, *Nature* **8**, 17 (2012)
  - [3] D.P. Varn, G.S. Canright, J.P. Crutchfield, *Acta Cryst. A* **69**, 197 (2013)
  - [4] V. Ryabov, D. Neroth, *Chaos* **21**, 037113 (2011)
  - [5] R. Haslinger, K.L. Klinker, C.R. Shalizi, *Neural computation* **22**, 121 (2010)
  - [6] D.P. Feldman. Computational mechanics of classical spin systems. (1998)
  - [7] D.P. Feldman, J.P. Crutchfield, Santa Fe Institute Working Paper 98-04-026 (2008)
  - [8] J.P. Crutchfield, *Physica D* **75**, 11 (1994)
  - [9] C.R. Shalizi, J.P. Crutchfield, *J. Stat. Phys.* **104**, 817 (2001)
  - [10] W. Löhr, *Adv. in Complex Syst.* **12**, 169 (2009)
  - [11] W. Löhr, *Entropy* **11**, 385 (2009)
  - [12] J.P. Crutchfield, D.P. Feldman, *Phys. Rev. E* **55**, R1239 (1997)
  - [13] D.P. Feldman, C.S. McTeque, J.P. Crutchfield, *Chaos* **18**, 043106 (2008)
  - [14] E. Behrends, *Introduction to Markov chain* (Springer, Heidelberg, 2000)
  - [15] P.A. Lavis, G.M. Dell, *Statistical mechanics of lattice systems 1: closed form and exact solutions* (Springer, Berlin, 1999)
  - [16] J.F. Dobson, *J. Math. Phys.* **10**, 40 (1969)
  - [17] J. Crutchfield, D.P. Feldman, *Chaos* **13**, 25 (2003)
  - [18] T. Morita, T. Horiguchi, *Physics Letters* **38a**, 223 (1972)
  - [19] N. Pottier, *Physica A* **230**, 563 (1996)
  - [20] R. Garcia-Pelayo, *Physica A* **384**, 143 (2007)
  - [21] Equation (7.2) in [6] is similar to equation (15), but the interpretation made of the parameters is different, as in [6] the product appearing in the numerator is identified with the product of the eigenvectors, which is not the case.
  - [22] This last equation together with eq. (25) shows that equation (7.6) and (7.15) in [6] are flawed.
  - [23] This result is equivalent to the one reported as equation (7.31) in [6], but, if one calculates all matrix entry from equation (7.15), the row normalization condition is violated. Therefore, the correctness of the entries in (7.31) in [6] is accidental as result of forcing row normalization.

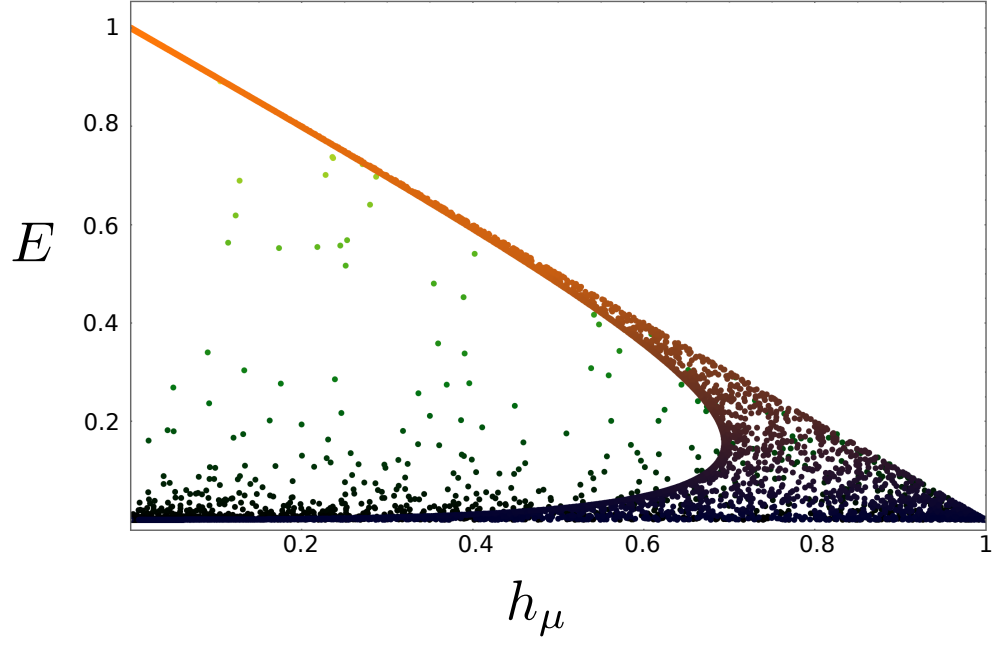


FIG. 1: The complexity-entropy diagram for the 1/2-nearest neighbor Ising model.  $10^5$  points were used with random parameters in the range  $\beta \in [10^{-4}, 10^2]$ ,  $J \in [-1.5, 1.5]$  and  $B \in [-3, 3]$ .

$$E_\mu(\beta \rightarrow \infty)$$

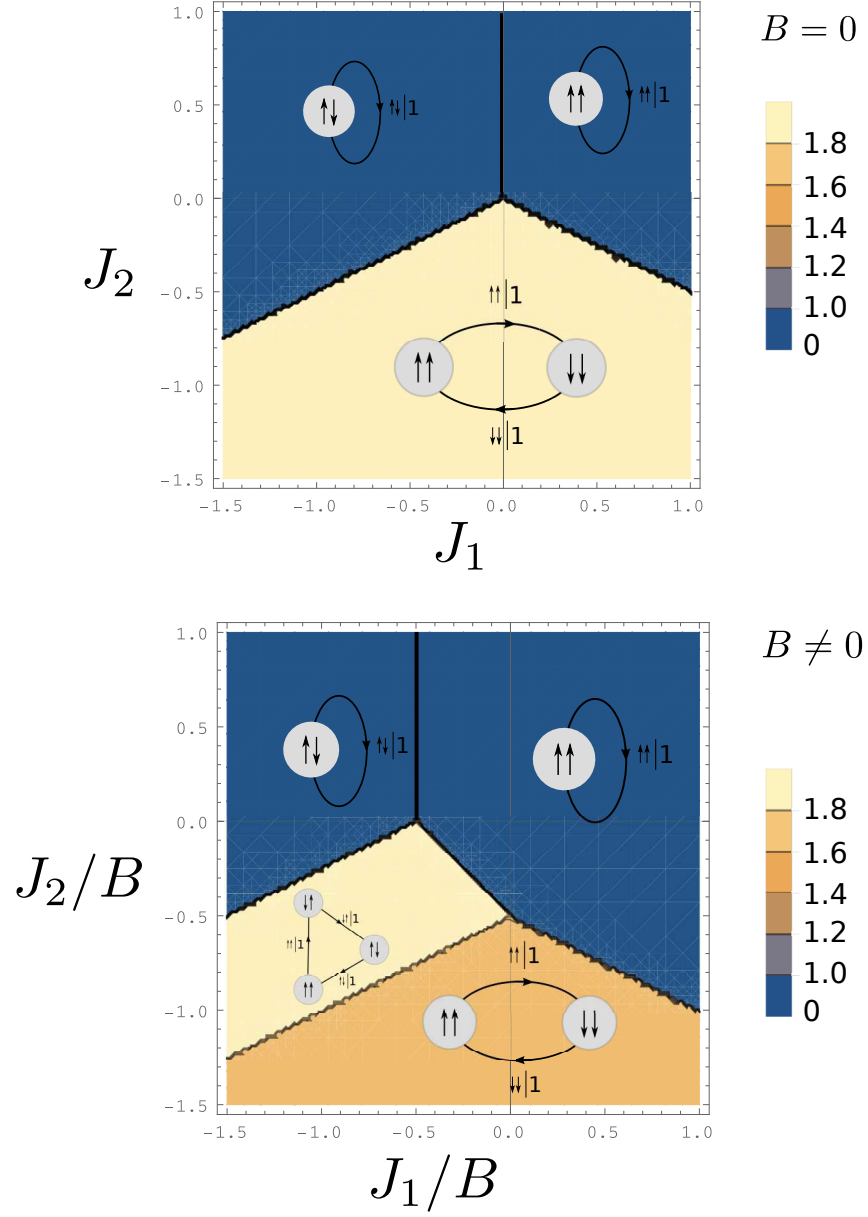


FIG. 2: Excess entropy ( $E_\mu$ ) at the ground state ( $\beta \rightarrow \infty$ ) for zero field ( $B = 0$ ) and non-zero field ( $B \neq 0$ ).

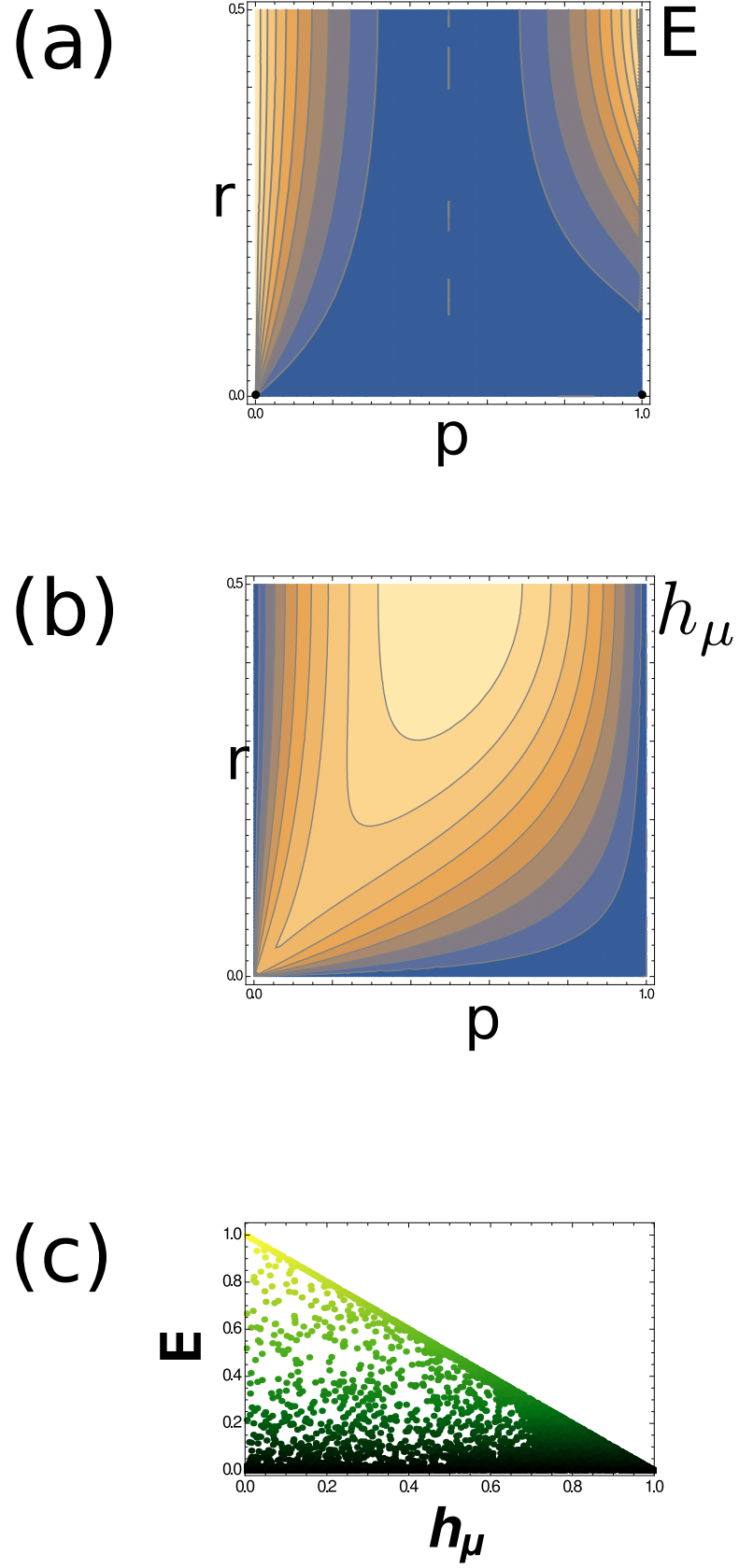


FIG. 3: Excess entropy (a) and entropy density (b) for different values of the probability of a right  $r$  movement and the probability of persistent  $p$  movement. (c) The complexity-entropy diagram of the persistent random walker,  $10^5$  points with random parameters  $(r, p)$  between 0 and 1 were used.



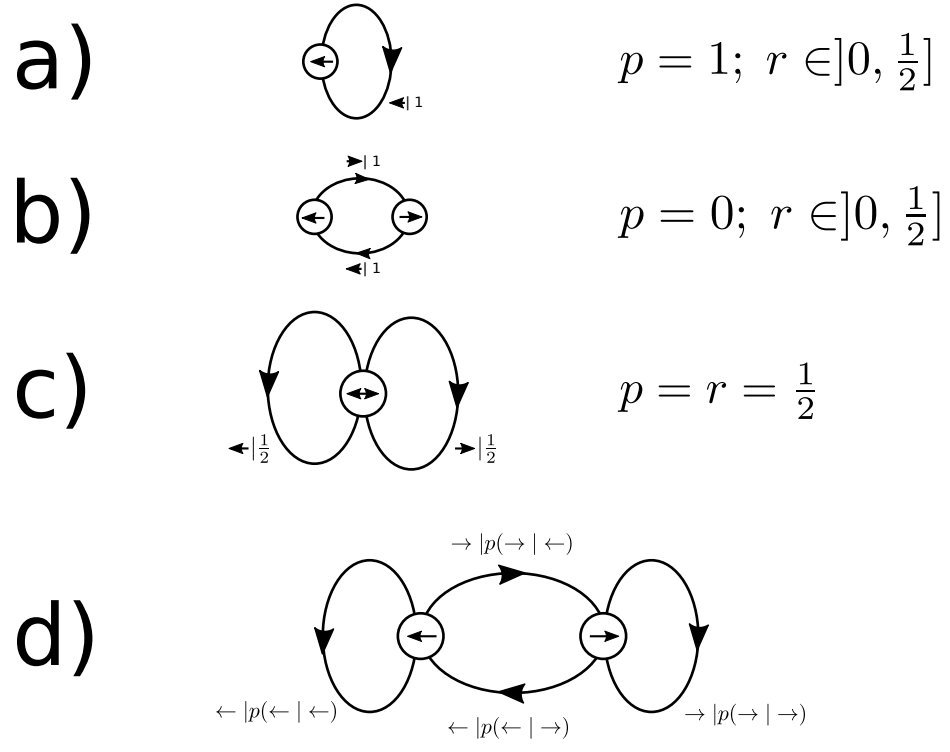


FIG. 4:  $\epsilon$ -machine diagrams as a finite state machine for the persistent random walker at different values of the “right” movement probability  $r$  and the persistence probability  $p$ .